Research Synopsis

"A Novel Scheme for marketing by study customer behavior using data mining technique."

Submitted By

Mr. Hambarde Kailas Anandrao

School of computational sciences

S.R.T.M.University,Nanded

Under the Guidance of

Dr. S. D. Khamitkar

Associate Professor,

School of Computational Sciences,

S. R. T. M. University, Nanded

Submitted to

Swami Ramanand Teerth Marathwada University, Nanded

1.Introduction:

In today'sworld most of the service provides and product companies while launching new products, services or releasing new version of existing product need to campaign to reach at the potential customers. While doing so they target their already customers who are the ambassadors of their company. To address the existing customers, they maintain the detailed customer data at all levels as customer maser data.

The majority of ambitious businesses maintain their customer data for future use. These data help them to establish and maintain a direct relationship with the customers in order to target them individually for specific product offers, services, opinion mining or fund raising in the future. Rather than targeting all the customers for all types of company's promotions, which does not result positive response, there is always an enthusiasm as who are the people would be targeted for the specific offer. In target marketing, it has been an arduous task to single out customers who are likely to be fascinated for a new product or service. At this juncture, data mining techniques can be applied for filtering out the target customers out of the pool. This would increase the overall effectiveness of the marketing campaign. Out of several standard techniques, selecting the best classification technique is a major task in data mining. As one size does not suit for all, likely one technique does not produce better yield for all types of data set. Therefore, few classification techniques have been applied on the data set. Furthermore, we have analyzed the result to obtain the most suitable technique.

Customer Segmentation: Sometimes referred to as market segmentation, customer segmentation is a method of analyzing a client base and grouping customers into categories or segments which share particular attributes. Key differentials in segmentation include age, gender, education, location, spending patterns and socio-economic group. Relevant differentials are those which are expected to influence customer behavior in relation to a business. The selected criteria are used to create customer segments with similar values, needs and wants.When planning a targeted marketing campaign, it is also necessary to differentiate customers within these groupings according to their preferred means of communication.

The Benefits of Segmentation

Customer segmentation is an essential tool in customer relationship management, enabling businesses to market effectively to their customers. Businesses are expected to

understand their customers and demonstrate their customer insights by sending only relevant, targeted communications to their customers. Customers want to feel valued and be treated as individuals, yet for anything other than perhaps the smallest of businesses this level of customer knowledge is impossible to achieve.

The main goal of this work is to study the existing data mining techniques and machine learning techniques used for customer segmentation. We propose new framework for customer segmentation, for that we generate customer statistical data. On this purposed framework we carry out a method to apply mostly used classification algorithms namely, support vector machine, logistic regression, K-Nearest Neighbors, Decision Tree, Random Forest and Gradient Boosting classifier.

Fromthe results, we can compare the effectiveness of the algorithmsfor the particular problem.Moreover, this is a technique to predict the customers whofrequently respond to theproduct offers based on the previoushistorical data.

2. Literature Review:

Targeting is the core of marketing management. It is concerned with offering the right product/service to the customer at the right time and using the proper channel. In this study author discussed how Data Mining modeling and analysis can support targeting applications. They focused on three types of targeting models: continuous-choice models, discrete-choice models and in-market timing models, discussing alternative modeling for each application and decision making. They also discussed a range of pitfalls that one needs to be aware of in implementing a data mining solution for a targeting problem.

- **Data Mining techniques for customer segmentation/ Target Marketing**

In this paper**[1]** author proposed a theoretical approach for data mining framework for target marketing. They provide a theoretical approach to gathering customer data and deriving relationship between the data so as to make marketing strategies effective. According to them this proposed system will help the analysts and various companies on basis of historical data to

predict who will respond to the marketing campaigns. In this study they used different data mining clustering methods such as Partitioning Method, Hierarchical Method, Density - based method and Grid - based methodfor customer segmentation

In this paper **[2]** author studied the DM techniques for credit card customer segmentation and target marketing. For this study author collated Chinese commercial bank's customer credit card data.This customer data classify into four classifications by K-means. After that they built forecasting model separately based on four data mining methods such as C5.0, neural network, chi-squared automatic interaction detector, and classification and regression tree according to the background information of the credit cards holders.

In this paper **[3**] author studied different data mining techniques used for customer segmentation by using a case study on Turkish supermarket chain. In this study, apriori algorithm is applied for mining association rules in database of Turkish supermarket chain.This database included customer information's like age, income and gender but without confidential information like name, address and phone. In this case study author found that the item which is sold the most was feta cheese, which was in 53 percent of all transactions. The second most sold item was tomato and it was present in 49 percent of transactions. These items were followed by, milk and pasta, respectively.

According to **[4]** conducting efficient and cost-effective targeted marketing using data mining techniques are more benefits. In this paper they have discussed bank term deposit prediction; also they investigate a few data mining algorithms and conducted experiments using these algorithms. The result implies that all the adopted data mining approaches can generate very high accurate rate. However, many of the algorithms generate high Type I and Type II error rates which is not ideal in the examined problem. High Type I error rate will require more human resources and time to contact the prospects. High Type II error rate signifies the adopted algorithm will miss real prospects that will bring revenue to the business. The experiment results also suggest that Naïve Bayes Multinomial can be the ideal algorithm to conduct efficient, cost-effective targeted marketing.

In this paper **[5]** author used clustering analysis for telecom customer segmentation. For analysis they used k-means clustering algorithm, and proposed the model process for telecom customer segmentation according to the principle of the algorithm. A resolution of customer segmentation for the Changzhou telecom in Jiangsu province is proposed, and practical results show that the clustering technology based on the K-means method provides an effective and successful resolution of customer segmentation for Telecom, bringing services closer to the customers.

In this paper **[6]** author constructs a three dimensional customer segmentation model based on customer lifetime value, customer satisfaction and customer activity, which more accurately divides customers into different groups. The corresponding variables are obtained by RFM model, Kano model and BG/NBD model. The customer segmentation model provides ten groups of customers with corresponding marketing strategies, so that it can help enterprises maximize profits.

In this study **[7]** author proposed a segmentation method that 1) divides the customer behavior sequences into cycles; 2) characterizes customers' cyclical behaviors based on the probability density distributions from the temporal dimension and frequency dimension, which could investigate the customer behaviors more comprehensively; 3) calculate the customer similarity by computing the difference of the distributions; 4) adopts the k-myeloid clustering algorithm to classify the customers based on the similarity matrix. The segmentation method is applied to a mobile online behavior dataset. The results of experiments indicate the relationship and typical customer usage regulations among various mobile e-commerce service groups, which will be informative for enterprises to understand their customers and improve their service quality

This paper**[8]**authors focused on the application of demand response (DR) in electricity customer classification and management, puts forward universal DR level index and load curve level index, constructing the customer classification index system. Principal component analysis (PCA) is used to reduce the original data dimension. Then the two methods for clustering are adopted and compared with the results.

In this research **[9]** article author tested 14 very different classification algorithms (random forest, gradient boosting machines, SVM - linear, polynomial, and RBF -1-hidden-layer neural nets, extreme learning machines, k-nearest neighbors and a bagging of knn, naive Bayes, learning vector quantization, elastic net logistic regression, sparse linear discriminate analysis, and a boosting of linear classifiers) on 115 real life binary datasets. They followed the Demas analysis and found that the three best classifiers (random forest, gbm and RBF SVM) are not significantly different from each other. They also discussed that a change of less than 0.0112 in the error rate should be considered as an irrelevant change, and used a Bayesian ANOVA analysis to conclude that with high probability the differences between these three classifiers is not of practical consequence. They also verified the execution time of \standard implementations" of these algorithms and concluded that RBF SVM is the fastest (significantly so) both in training time and in training plus testing time.

In this research **[10]** authors measuring net sentiment based on customer satisfaction through customer's sentiment analysis from Twitter data. They developed sample model is built and extracted from more than 3.000 raw Twitter messages data from March to April 2016 of top marketplace in Indonesia. They compared several algorithms, and the classification schemes. The sentiments are classified and compared using five different algorithms classification. There are, K-Nearest Neighbor, Logistic Regression, Naïve Bayes, Random Forest, and Support Vector Machine. The five machine learning can be applied to the Indonesian-language sentiment analysis. Preprocessing on thes tages of tokenization, parsing, and stop word deletion of word frequency counting. Frequency of the word of the document used weighted by TF-IDF. The Random Forest, Support Vector Machine, and Logistic Regression generate better accuracy and stable compared with the Naïve Bayes and K-Nearest Neighbor. The results showed Support Vector Machine has accuracy81.82% with 1000 sampling dataset and 85.4% with 2000 sampling dataset. This shows that the more the number of training data will improve the accuracy of the system. The Net Sentiment score for marketplace in Indonesia is 73%. This results also showed that customer satisfaction has average Net Promoter Score (NPS) 3.3%.

In this paper [11] authors had integrated Customer Relationship Management and Data Mining Framework for Customer Classification and Risk Analysis in Health Sector In this study, the data of a company which is operating in health sector in Turkey was analyzed using some

data mining algorithms like Naive Bayes which is an algorithm of classification rule, J48 which is one of the algorithms to generate decision tree and k-nearest neighbour IBk algorithm.

This [12] study authors investigated the influence of problems with data accuracy – an important dimension of data quality – on three prominent segmentation techniques for direct marketing: RFM (recency, frequency, and monetary value) analysis, logistic regression, and decision trees. For two real-life direct marketing data sets analyzed, the results demonstrate that (1) under optimal data accuracy, decision trees are preferred over RFM analysis and logistic regression; (2) the introduction of data accuracy problems deteriorates the performance of all three segmentation techniques; and (3) as data becomes less accurate, decision trees retain superior to logistic regression and RFM analysis. Overall, this study recommends the use of decision trees in the context of customer segmentation for direct marketing, even under the suspicion of data accuracy problems.

According to [**13**] most of the service provides and product companies while launching name new products, services or releasing new version of existing product need to campaign to reach at the potential customers. While doing so they target their already customers who are the ambassadors of their company. To address the existing customers, they maintain the detailed customer data at all levels as customer maser data. For addressing this issue author had built a prediction model to identify the customers who would most likely respond to the prospective offerings of the company basing on their past purchasing trends. They conducted experiments using the well know classifiers, viz, Navin Bayes, KNN and SVM to classify a bank customer data. In this study author found that among the all algorithm that Naïve Bayesian Classification achieves the highest accuracy and specificity. However, the worst classification was performed by support vector machine as SVM technique achieves lowest sensitivity, accuracy and specificity.

3. Research Outline

This section will examine the research question, aim, objectives, scope, significance, limitations and the justifications for this study.

3.1Research Question

The research question examined by this thesis is:

- How the New Emerging Machine Learning (ML) Technology and Existing Data Mining (DM) plays role in Customer Recommendation System or customer segregation based on their previous pursing pattern?

It hypothesized that a Machine Learning based decision support system as well as from the collected dataset a pattern can be extracted by applying various clustering algorithm to build a decision support system for target marketing.

3.2 Aim

The aim of this research is to develop predictive model based on customer's previous pursing behavior using machine learning and Data Mining techniques for Customer segmentation as well as target marketing.

3.3 Objectives

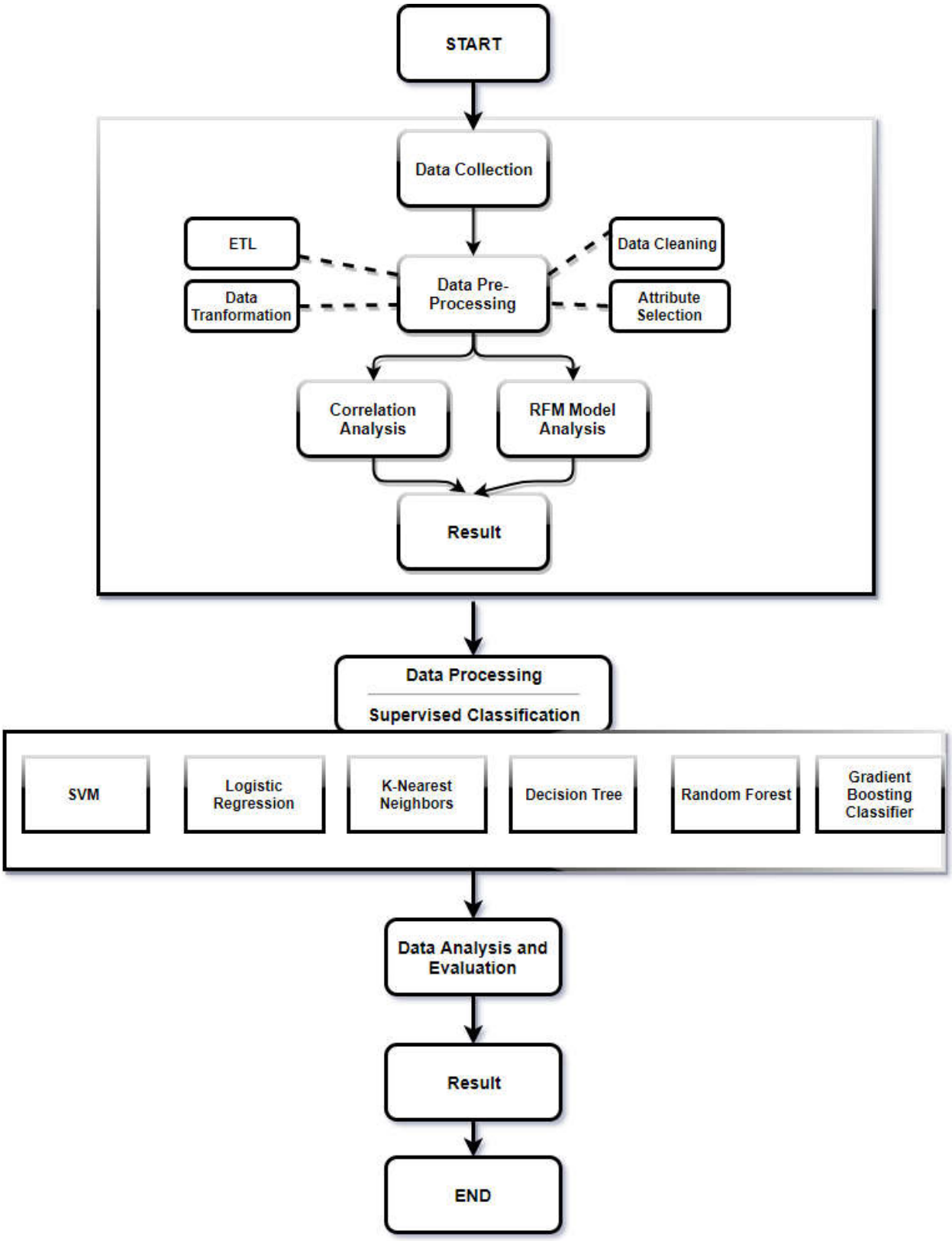
1) To analyze the existing techniques of DMT and MLT for customer classification and segmentation.
2) Generate customer's statistical data for purposed framework.
3) Study the different unsupervised machine learning algorithm used for customer segmentation.
4) Study the supervised machine learning classification algorithms for customer classification.
5) Develop statistical model for customer segmentation.

6) Develop predication model to identify the customer who would mostly likely respond to prospective.
7) Apply machine learning classification algorithm to purposed model.
8) Study of best supervised classification algorithms with accuracy assessment.
    1. Support Vector Machine Classifier (SVC)
        a. Confusion matrix
        b. Learning curves
    2. Logistic regression
    3. k-Nearest Neighbors
    4. Decision Tree
    5. Random Forest
    6. Gradient Boosting Classifier
9) Comparative analysis of supervised classification algorithms.

3.4 Significance

- The major contribution of this research can be summarized as:
- It gives factual decision support system to business owner for customer classification.
- Develop new Predicative model for target marketing using statistical and machine learning techniques.

5. Methodology:



6. Expected Outcome

Focused on the methodology, this study aims to use Data Mining and Machine Learning Algorithm to mine knowledge from various data and integrate previous mine knowledge to analyze and predict the best customer for target marketing. The outcome from overall juncture lead to minor enhancements of the original theoretical outline and also provides new research question for future work.

7. Reference:

1. Ritika Pandey *"TARGET MARKETING USING DATA MINING"* International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 Volume-4, Issue-6, Jun-2017.
2. Li, W., Wu, X., Sun, Y., & Zhang, Q. (2010, December). Credit card customer segmentation and target marketing based on data mining. In *Computational Intelligence and Security (CIS), 2010 International Conference on* (pp. 73-76). IEEE.
3. Güllüoğlu, S. S. (2015, February). Segmenting customers with data mining techniques. In *Digital Information, Networking, and Wireless Communications (DINWC), 2015 Third International Conference on* (pp. 154-159). IEEE.
4. Liu, L., Yang, Z., & Benslimane, Y. (2013, December). Conducting efficient and cost-effective targeted marketing using data mining techniques. In *Intelligent Systems (GCIS), 2013 Fourth Global Congress on* (pp. 102-106). IEEE.
5. Qiuru, C., Ye, L., Haixu, X., Yijun, L., & Guangping, Z. (2012, August). Telecom customer segmentation based on cluster analysis. In *Computer Science and Information Processing (CSIP), 2012 International Conference on* (pp. 1179-1182). IEEE.
6. He, X., & Li, C. (2016, December). The Research and Application of Customer Segmentation on E-Commerce Websites. In *Digital Home (ICDH), 2016 6th International Conference on* (pp. 203-208). IEEE.
7. Zhao, H., Zhang, X. H., Wang, Q., Zhang, Z. C., & Wang, C. Y. (2014, August). Customer segmentation on mobile online behavior. In *Management Science & Engineering (ICMSE), 2014 International Conference on* (pp. 103-109). IEEE.
8. Wang, P., & Zhang, P. (2017, June). Classification and management of electricity market customer considering demand response in China. In *Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), 2017 IEEE International Conference on* (pp. 1-5). IEEE.

9. Wainer, J. (2016). Comparison of 14 different families of classification algorithms on 115 binary datasets. *arXiv preprint arXiv:1606.00930*.

10. Wainer, J. (2016). Comparison of 14 different families of classification algorithms on 115 binary datasets. *arXiv preprint arXiv:1606.00930*.

11. Alsaç, A., Çolak, M., & Keskin, G. A. (2017, March). An integrated customer relationship management and Data Mining framework for customer classification and risk analysis in health sector. In *Industrial Technology and Management (ICITM), International Conference on* (pp. 41-46). IEEE.

12. Coussement, K., Van den Bossche, F. A., & De Bock, K. W. (2014). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of Business Research*, *67*(1), 2751-2758.

13. Das, T. K. (2015, October). A customer classification prediction model based on machine learning techniques. In *Applied and Theoretical Computing and Communication Technology (iCATccT), 2015 International Conference on* (pp. 321-326). IEEE.

14. Kandeil, D. A., Saad, A. A., & Youssef, S. M. (2014, September). A Two-Phase Clustering Analysis for B2B Customer Segmentation. In *Intelligent Networking and Collaborative Systems (INCoS), 2014 International Conference on* (pp. 221-228). IEEE.